Emerging single cell tools are primed to reveal functional and molecular heterogeneity in malignant hematopoietic stem cells

Mairi S. Shepherd[1,2] and David G Kent[1,2]

[1]Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, Hills Road, Cambridge, CB2 0XY, United Kingdom

[2]Department of Haematology, University of Cambridge, CB2 0XY, United Kingdom

Address correspondence:

David G. Kent, Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, CB2 0AH, United Kingdom

dgk23@cam.ac.uk

Tel: +44 1223 762130

Abstract

Purpose of Review

The recent emergence of single cell technologies has permitted unprecedented insight into the molecular drivers of fate choice in blood stem and progenitor cells. This review gives a broad overview of current efforts to understand the molecular regulators of malignant HSCs at the single cell level.

Recent Findings

The large-scale adoption of single cell approaches has allowed extensive description of the transcriptional profiles and functional properties of single HSCs. These techniques are now beginning to be applied to malignant HSCs isolated directly from patients or from mouse models of malignancy. However, these studies have generally struggled to pinpoint the functional regulators of malignant characteristics, since malignant HSCs often differ in more than one property when compared to normal HSCs. Moreover, both normal and malignant populations are complicated by HSC heterogeneity.

Summary

Despite the existence of single cell gene expression profiling tools, relatively few publications have emerged. Here, we review these studies from recent years with a specific focus on those undertaking single cell measurements in malignant stem and progenitor cells. We anticipate this to be the tip of the iceberg, expecting the next 2-3 years to produce datasets that will facilitate a much broader understanding of malignant HSCs.

Key words: Single-cell, HSC, malignancy, heterogeneity, gene expression

Introduction

Over the last 5 years, there has been an explosion of new tools to study the biology of large numbers of single cells[1,2]. Functional cell biology techniques have long been used to study single cells, but new molecular techniques have allowed the coupling of cellular and molecular heterogeneity[3–6], adding unprecedented resolution to the concepts first investigated using cellular biology techniques.

Functional heterogeneity in HSCs

The existence of heterogeneity in hematopoietic stem cells (HSCs) has been well documented from a functional standpoint. In the 1960s, Becker, McCulloch and Till first alluded to the cellular heterogeneity of HSCs in colony forming unit spleen (CFU-S) assays[7], a seminal *in vivo* assay which demonstrated that the output of hematopoietic progenitors varied considerably in their size, composition (the numbers and types of mature and primitive cells), time to detection and the number of secondary colonies they could give rise to. These differences were the first hint that heterogeneity was a feature of mammalian blood stem cells and came at a time when bone marrow transplantations were emerging as a clinical option[8,9]. Key experiments in the 1980s tracking HSCs using retroviral inserts as genetic marks for their progeny, provided the first formal evidence of variations in self-renewal durability and also reaffirmed differences in mature cell production, prompting the eventual description of lineage-biased HSCs[10–13]. Perhaps the greatest advance in deciphering the functional heterogeneity of HSCs was the introduction of single cell florescence activated cell sorting (FACS) for HSC purification[14], which lead to single HSC sorting and transplantation assays[15].

The gold-standard assay for evaluating self-renewal is serial transplantation into myeloablated recipient mice[16]. Prior to the advent of single cell transplantations, limiting dilution assays were used to approximate the number of HSCs within a test cell suspension. Results were commonly quantified using the competitive repopulating unit (CRU)[17] since it could not be formally determined that a single cell gave rise to the donor graft. This uncertainty was alleviated when single cell transplantations formally demonstrated that one cell could on its own be isolated and transplanted into recipients to give rise to long-term, multi-lineage haematopoiesis[15]. This remains the best means for assessing HSC function with respect to lineage output, expansion capacity and long-term self-renewal durability.

However, single cell transplantations have several limitations. The first of these is that it assesses what an HSC *can* do in a stressed environment, as opposed to what it *does* do under homeostatic conditions. The second is that it does not test how HSCs interact as a population to maintain the balance of cellular outputs. These issues are at least partially overcome by genetic barcoding studies where bulk HSPCs can be transplanted together and tracked individually[18]. However, these assays are still transplantation-based and are further limited by needing an *in vitro* culture period, meaning that the culture itself (or integration site of the barcode) could add bias to the population. Moreover, sequence detection limits could also add bias, as it cannot conclusively determine absence of contribution or disappearance of a clone in longitudinal studies. More advanced endogenous barcoding approaches have now been undertaken where HSCs are traced using reporter mice[19] or genetic recombination is used to mark cells *in vivo* without any additional manipulation[1,19,20] .

It is clear from the above studies that understanding HSC heterogeneity has been a longstanding goal of the field, likely due to its implications for understanding the evolution of the immune system and the development of aberrant haematopoiesis, e.g. age-related clonal haematopoiesis (ARCH) and leukemia. HSC subtypes have been described in mice based on differences in their self-renewal durability and mature cell production[21–24], and this has prompted efforts to prospectively identify individual subtypes[25–27]. Cumulatively, efforts have produced overall HSC purities of >50% and several enrichment strategies for enriching various subtypes, setting the stage for studies to define the molecular programme governing HSC subtypes (both normal and malignant). This increased ability to purify HSCs and the concomitant development of robust global gene expression tools in single cells[28] offer an obvious solution to this longstanding problem. This review summarises recent efforts in this space.

Molecular heterogeneity in HSCs

Single cell molecular profiling of HSC and progenitors has already revealed a number of findings that were previously unattainable using bulk populations. Unbiased approaches such as RNAseq have identified new genes (and associated pathways) involved in stem cell function and lineage commitment[29,30]. These same techniques unveiled the heterogeneity within various progenitor compartments[30], and confirmed that the molecular process by which stem cells differentiate is a gradual process (i.e., a continuum) rather than a stepwise progression though progressively more differentiated progenitors[31]. Importantly, studies that have profiled HSCs all converge on the

description of significant heterogeneity within the population, but the utility of such a finding is unclear.

An important factor when considering "heterogeneity" is the difference between informative heterogeneity (biologically driven variations that result in functional differences), and generic heterogeneity (differences driven by processes occurring in all cell types (e.g., cell cycle[28]). Another consideration is the extent of heterogeneity that is actually driven by technical noise: if it were possible to sequence the same cell many times, there would be technical drop-out of genes and this would differ between experiments, adding an artificial "heterogeneity" to the population. Finally, and perhaps least obvious, there is yet another type of heterogeneity hidden amongst molecular studies of purified HSCs – that of non-HSC contaminants. Unlike transplantation assays, where non-HSCs do not read out and are therefore not able to confuse the description of heterogeneity, molecular assays will generate 100 gene expression profiles from 100 cells. When populations are of a low purity eg 5 or 10% and single cell molecular studies are undertaken, it becomes incredibly difficult to assign a specific molecular programme to a specific cell function. The reality in such cases is that 90-95% of the molecular programmes are NOT the cell of interest and, unlike long term cell biological assays where non-HSCs are no longer represented, these represent a significant confounder for single cell molecular biology studies (Figure 1).

The latter issue can be partially circumvented by linking functional single cell HSC assays to molecular assays. Flow cytometric index sorting records the fluorescence intensity of each immunophenotypic marker, thereby allowing the retrospective coupling of markers and output. This has been combined with single cell functional *in vitro*[3] and *in vivo*[4] assays to refine the effectiveness of sort panels. Such approaches are already starting to demonstrate their utility in the human setting as well with *in vitro* functional readouts of human HSCs being linked to molecular profiles by index-sorting[32]. We predict that the next several years will yield many similar studies, but an additional challenge looms when trying to understand how single HSCs are corrupted from a molecular standpoint to drive disease.

HSC heterogeneity and malignancy

Malignant HSCs share many of the same cellular processes and pathways as normal HSCs, but some processes must be hijacked to enable increased proliferation and/or a differentiation block. Therefore, it becomes important to not simply think of malignant

HSCs as a completely separate cell state compared to normal HSCs, but rather a very close molecular relative. A reductive approach leads many researchers to study highly purified populations of HSCs in mouse models bearing a single genetic change. Since many myeloid malignancies have now been extensively sequenced and numerous malignancy-driving mutations identified, a great deal of effort has focused on establishing the mechanism by which these mutations initiate and sustain disease.

It is in studying the molecular effects of specific mutations on HSCs that mouse models are particularly useful, largely because we can isolate HSCs at a much higher frequency compared to human, and additionally mouse models can be 100% mutant, thereby avoiding complications of intra-patient HSC heterogeneity in mutational state[33]. From mouse models we have determined that there are common pathways driving shared disease features, (e.g., hyperactivation of the JAK/STAT signalling pathway leads to a proliferative advantage and can be achieved though mutations in JAK1,2, STAT1,3,5, MPL, CALR, LNK and CBL)[34].

Taking advantage of these mouse models, many studies have now analysed the transcriptional profiles of mutant mouse models at the bulk level and it has become clear that progress is impeded by two main issues: 1) population heterogeneity (not purifying the HSC fraction) and 2) an inability to observe co-expression (or not) in the same cell. Single cell approaches give this added resolution when used appropriately, although the number of published transcriptional profiles in mouse models of hematological malignancies at the single cell level remains limited (Table 1).

Since haematological malignancies often have more than one driver[35], and several aspects of HSCs can be affected by each mutation (e.g., JAK2 mutations alter cell cycle status, proliferation, differentiation and HSC self-renewal), it remains difficult to ascribe individual properties of disease when studying single mutations in isolation. For these reasons, many recent studies have generated mouse models to study combinations of mutations[36–42].

However as with single mutant models, there is a paucity of single cell molecular assays carried out in these models. In these limited instances, it is clear that single cell approaches can make novel discoveries that are not possible to uncover by studying single mutations alone. To cite one example, in a study of a mouse model of CN-AML[43], where DNMT3a and $FLT3^{ITD}$ mutations were combined, single cell RNA-seq was

performed on c-Kit+ leukemic splenocytes to determine whether c-Myc expression could be a biomarker for FLT3^{ITD} (it could not). The heterogeneity of AML tumors, coupled with incomplete penetrance of the mutations, made this question unanswerable at the bulk level, demonstrating the power of single cell approaches to reveal new biology.

Integrating functional and molecular heterogeneity in malignant HSCs

Combining the latest advances in single cell techniques with recently developed mouse models, it is now possible to understand more deeply the precise mechanisms involved in malignant HSC self-renewal. To undertake such studies, simplified disease models are useful. Myeloproliferative neoplasms (MPNs) are a particularly tractable disease for studying leukemogenesis because they are a relatively simple disease with a low mutational burden (the vast majority of patients have fewer than 3 mutations[44]). The most common mutation in MPNs is JAK2 V617F and a number of mouse models have been developed to understand how one mutation can be associated with multiple disease subtypes[45], ARCH[46,47], or potentially in undetectable transient clones. These mouse models all recapitulate some hematological aspects of disease (e.g., raised RBCs or platelets), but - like MPNs in patients - differ from each other in several respects. The first gene expression profiles of JAK2 V617F HSCs were performed at the bulk level and the only changes that could be observed (if any) pointed to differences in cell cycle regulators[48], which is only one of the properties altered in malignant HSCs.

Single cell approaches in combinatorial mouse models can help resolve this issue. As it stands, it is impossible to know *a priori* which molecules are involved with each altered property in a single-mutant mouse model (e.g., hyper-proliferation, dormancy, HSC self-renewal changes or differentiation patterns). We recently tried to address these issues by crossing JAK2 V617F mice with TET2 KO mice[48]** (TET2 is an epigenetic regulator that is the most common co-mutation with JAK2 V617F). Since TET2 KO mice have no erythrocytic phenotype and display an HSC self-renewal advantage, the comparison of single-mutant and double-mutant mice allowed us to study 4 unique states: normal unperturbed WT HSCs, hyper-proliferating HSCs with a self-renewal defect (JAK2V671F single-mutant HSCs), HSCs with a self-renewal advantage and relatively normal proliferation/differentiation (TET2 single-mutant HSCs), and JAK2/TET2 double mutant HSCs. The first observation we made was that the double-mutant HSCs represented some, but not all aspects of each mutations – HSCs had normal self-renewal but remained hyper-proliferative and their progeny had the differentiation abnormalities associated with the JAK2V617F mutation.

This series of models permitted us to ask a direct question about the regulation of malignant HSC self-renewal: which genes are responsible for the JAK2 V617F self-renewal defect? We answered this by profiling 43 established HSC regulators, in combination with single cell functional assays in each of these models. This revealed new candidate molecules from which *Bmi1*, *Pbx1* and *Meis1* appeared to have functional consequences on malignant HSC self-renewal. Thus our study exemplified how single cell approaches, in combination with allelic series, can be used to unveil new biological information not possible with bulk studies. Applying similar approaches more broadly across new models will help us understand the molecular networks driving each individual property of malignant HSCs.

Human patient HSCs

The goal of understanding the molecular networks governing malignant HSC self-renewal is to apply the findings to treating patients with cancer. It is therefore desirable for similar molecular studies to be performed on patient samples. Such studies are limited, however, owing to a number of factors: firstly, as mentioned above the efficiency of sorting functional human HSCs using immunophenotypic markers is much lower than in mouse (<10% in human >50% in mouse), which has obvious consequences for downstream gene expression analyses. Secondly, disease burden is variable in patient samples and currently there are no markers in hematological malignancies for prospectively identifying mutant versus non-mutant cells meaning that samples would be different mixtures of malignant and non-malignant HSCs. Thirdly, patients are extremely heterogeneous and vary in a number of aspects related to disease such as the combinations of mutations they have[49], the order of mutation acquisition[50], and differences in lifestyle; smoking, diet, etc[51]. These factors each add a layer of complexity to the study of the human HSCs and their molecular networks.

One of the first studies to try and tackle this latter issue took a clonal approach to distinguish mutant and non-mutant clones retrospectively[33]. Growing and genotyping erythroid colonies from a single patient enabled transcriptional profiling of non-mutant, heterozygous JAK2V617F and homozygous JAK2V617F cells, thereby controlling for intra-patient heterogeneity. One of the most important revelations of this paper, however, was that across 36 patients studied in this way, colony transcriptomes clustered much more strongly within the patient than with the individual mutations, confirming that intra-patient heterogeneity is a substantial confounding variable in such studies. Also, transcriptomes could only be collected from the differentiated progeny of HSCs

rather than the HSCs themselves, so only so much could be learned about the HSC transcriptome itself.

To try and address the issue of directly measuring a mutant HSC transcriptome, developments in single cell molecular biology again proved useful. The most impressive effort to date coupled single cell transcriptomics with detection of the BCR-ABL transgene in chronic myeloid leukemia (CML) patients by multi-plexing BCR-ABL-specific primers at the reverse transcription and amplification steps of the scRNA-seq protocol[2]**. This technology allows sensitive and specific detection of the BCR-ABL mutation simultaneously with unbiased whole transcriptome analysis of the same HSC, thereby permitting a molecular comparison of mutant to non-mutant HSCs in the same individual. After validating the nested priming approach for genotyping cells from the transcriptome, this study compared HSCs from normal individuals to BCR-ABL+ HSCs and BCR-ABL- HSCs from CML patients.

Perhaps the most interesting aspect of the Giustacchini study was their comparison of normal HSCs to non-mutant HSCs in the CML patient which showed striking differences in the gene signatures. Non-mutant HSCs from CML patients had higher expression of genes associated with microenvironmental factors IL-6, STAT5, TGF-β and TNF-α. As inflammation is a suspected suppressor of HSC function[52,53], this suggests that in humans, leukemic stem cells and their progeny might be creating their own self-supporting niche[54] that suppresses normal non-mutant HSCs. Moreover, response of patients to TKI treatment could be predicted by the inflammatory signalling changes observed in non-mutant HSCs – again something that could only be detected by being able to study these HSCs separately.

However BCR-ABL, unlike other mutations, only requires a single genotyping assay to be developed for all patients, loss-of-function mutations would be less straightforward to genotype via the transcriptome in a robust and scalable manner. Efforts to scale this to more mutations have nevertheless begun in diseases such as AML[55] but the efficiencies are not yet at a point where there is a high confidence in calling the absence of a mutation. To address this issue Rodriguez-Meira et al[56]* have developed a technique combining scRNA-seq with targeted mutation sequencing from gDNA and cDNA, allowing much more accurate calling of mutations for which cDNA expression is undetectable or highly allelic-biased. If such approaches were broadly applicable across many mutations in a

reliable and quantitative fashion, this would revolutionise the field of patient HSC analysis.

Key points:

- Functional heterogeneity of HSCs has been well established, first alluded to in the 1960s and robustly supported by single cell functional assays
- Single cell gene expression assays are trendy but biological, generic and contaminating heterogeneity must be considered when interpreting
- Groups are now combining functional and molecular techniques to study mutation combinations and malignant HSC properties in mouse and human.
- The stage is set for an explosion of papers using single-cell techniques to investigate haematological malignancies

Concluding thoughts

Single cell approaches are leading the way to a much more profound understanding of the molecular networks governing malignant HSC fate choice. New mouse models are constantly improving our understanding of the common pathways driving malignancy, but a gap remains in our understanding around the interplay between these single mutations and additional genetic and non-genetic factors. Many patients have combinations of mutations, and while several groups have started modelling these interactions[36–42], very few have studied the molecular biology of HSCs at the single cell level[48].

Studies in recent years have demonstrated that the hematopoietic microenvironment can have a substantial impact upon the cellular output of normal and malignant HSCs[54]. Perturbation of the hematopoietic system by a number of different exogenous insults likely impacts the growth and expansion of HSC clones with different mutations. Recent evidence shows that infections can affect the HSCs directly[57–60].

Beyond understanding the molecular nature of malignancy, we envisage a growing need to monitor differential HSC clonal dynamics in humans to understand how a single clone can emerge to drive a leukaemia. Recent evidence suggests that using somatic mutation acquisition as an endogenous barcode might be useful for studying the relative contribution of clones over the lifetime of an individual[61]*. At present, such techniques are not possible in large numbers of humans, but if similar approaches could be devised, studying how clonal dynamics change upon acquisition of oncogenic mutations and

exposure to environmental or endogenous stresses may well become a fruitful line of enquiry. Until then, *in vitro* approaches and mouse models will be heavily relied upon. The emergence of these new single cell and clonal technologies has generated significant enthusiasm amongst researchers and clinicians trying to understand the molecular differences between malignant and non-malignant cells, including a more complete understanding of clonal competition during disease establishment, maintenance and progression. As mentioned above, the first studies are only now starting to emerge as technologies converge; the next three to five years promise to deliver an increasingly detailed understanding of the molecules governing fate choice in normal and malignant HSCs.

1. Lyne, A.-M. *et al.* A track of the clones: new developments in cellular barcoding. *Exp. Hematol.* **68**, 15–20 (2018).

** 2. Giustacchini, A. *et al.* Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat. Med.* **23**, 692–702 (2017).

Using novel technology to distinguish mutant from non-mutant cells in patients with CML, allowing conclusions to be drawn on the effect of cancer on non-mutant cells within the patients.

3. Schulte, R. *et al.* Index sorting resolves heterogeneous murine hematopoietic stem cell populations. *Exp. Hematol.* **43**, 803–811 (2015).
4. Wilson, N. K. *et al.* Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell* **16**, 712–724 (2015).
5. Dahlin, J. S. *et al.* A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. *Blood* **131**, (2018).
6. Psaila, B. *et al.* Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. *Genome Biol.* **17**, 83 (2016).
7. Becker, A. J., McCulloch, E. A. & Till, J. E. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197**, 452–4 (1963).
8. Eaves, C. J. Hematopoietic stem cells: concepts, definitions, and the new reality. *Blood* **125**, 2605–13 (2015).
9. Muller-Sieburg, C. E., Sieburg, H. B., Bernitz, J. M. & Cattarossi, G. Stem cell

heterogeneity: implications for aging and regenerative medicine. *Blood* **119**, 3900–3907 (2012).

10. Lemischka, I. R., Raulet, D. H. & Mulligan, R. C. Developmental potential and dynamic behavior of hematopoietic stem cells. *Cell* **45**, 917–27 (1986).
11. Jordan, C. T. & Lemischka, I. R. Clonal and systemic analysis of long-term hematopoiesis in the mouse. *Genes Dev.* **4**, 220–32 (1990).
12. Dick, J. E., Magli, M. C., Huszar, D., Phillips, R. A. & Bernstein, A. *Introduction of a Selectable Gene into Primitive Stem Cells Capable of Long-Term Reconstitution of the Hemopoietic System of W/W v Mice*. *Cell* **42**, (1985).
13. Keller, G. & Snodgrass, R. Life span of multipotential hematopoietic stem cells in vivo. *J. Exp. Med.* **171**, 1407–18 (1990).
14. Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58–62 (1988).
15. Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-Term Lymphohematopoietic Reconstitution by a Single CD34-Low/Negative Hematopoietic Stem Cell. *Science (80-.).* **273**, 242–245 (1996).
16. Kent, D. G., Dykstra, B. J. & Eaves, C. J. Isolation and Assessment of Single Long-Term Reconstituting Hematopoietic Stem Cells from Adult Mouse Bone Marrow. *Curr. Protoc. Stem Cell Biol.* **38**, 2A.4.1-2A.4.24 (2016).
17. Szilvassy, S. J., Humphries, R. K., Lansdorp, P. M., Eaves, A. C. & Eaves, C. J. Quantitative assay for totipotent reconstituting hematopoietic stem cells by a competitive repopulation strategy. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8736–40 (1990).
18. Naik, S. H. *et al.* Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature* **496**, 229–232 (2013).
19. Pei, W. *et al.* Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* **548**, 456–460 (2017).
20. Rodriguez-Fraticelli, A. E. *et al.* Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212–216 (2018).
21. Dykstra, B. *et al.* Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell Stem Cell* **1**, 218–29 (2007).
22. Ema, H., Morita, Y. & Suda, T. Heterogeneity and hierarchy of hematopoietic stem cells. *Exp. Hematol.* **42**, 74–82.e2 (2014).
23. Pinho, S. *et al.* Lineage-Biased Hematopoietic Stem Cells Are Regulated by Distinct Niches. *Dev. Cell* **44**, 634–641.e4 (2018).
24. Yamamoto, R. *et al.* Clonal Analysis Unveils Self-Renewing Lineage-Restricted

Progenitors Generated Directly from Hematopoietic Stem Cells. *Cell* **154**, 1112–1126 (2013).

25. Morita, Y., Ema, H. & Nakauchi, H. Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment. *J. Exp. Med.* **207**, 1173–82 (2010).
26. Beerman, I. *et al.* Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *Proc. Natl. Acad. Sci.* **107**, 5465–5470 (2010).
27. Kent, D. G. *et al.* Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential. *Blood* **113**, 6342–50 (2009).
28. Wilson, N. K. & Göttgens, B. Single-Cell Sequencing in Normal and Malignant Hematopoiesis. *Hemasphere* **2**, (2018).
29. Paul, F. *et al.* Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* **163**, 1663–1677 (2015).
30. Olsson, A. *et al.* Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature* **537**, 698–702 (2016).
31. Laurenti, E. & Göttgens, B. From haematopoietic stem cells to complex differentiation landscapes. *Nat. 2018 5537689* **553**, 418 (2018).
32. Belluschi, S. *et al.* Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors. *Nat. Commun.* **9**, 4100 (2018).
33. Chen, E. *et al.* Distinct Clinical Phenotypes Associated with JAK2V617F Reflect Differential STAT1 Signaling. *Cancer Cell* **18**, 524–535 (2010).
34. Thomas, S. J., Snowden, J. A., Zeidler, M. P. & Danson, S. J. The role of JAK/STAT signalling in the pathogenesis, prognosis and treatment of solid tumours. *Br. J. Cancer* **113**, 365 (2015).
35. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–74 (2013).
36. Guryanova, O. A. *et al.* DNMT3A mutations promote anthracycline resistance in acute myeloid leukemia via impaired nucleosome remodeling. *Nat. Med.* **22**, 1488–1495 (2016).
37. Obeng, E. A. *et al.* Physiologic Expression of Sf3b1 K700E Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. *Cancer Cell* **30**, 404–417 (2016).

38. Shih, A. H. *et al.* Mutational Cooperativity Linked to Combinatorial Epigenetic Gain of Function in Acute Myeloid Leukemia. *Cancer Cell* **27**, 502–515 (2015).
39. Shimizu, T. *et al.* Loss of Ezh2 synergizes with JAK2-V617F in initiating myeloproliferative neoplasms and promoting myelofibrosis. *J. Exp. Med.* **213**, 1479–96 (2016).
40. Yang, L. *et al.* DNMT3A Loss Drives Enhancer Hypomethylation in FLT3-ITD-Associated Leukemias. *Cancer Cell* **29**, 922–934 (2016).
41. Chen, E. *et al.* Distinct effects of concomitant Jak2V617F expression and Tet2 loss in mice promote disease progression in myeloproliferative neoplasms. *Blood* **125**, 327–335 (2015).
42. Loberg, M. A. *et al.* Sequentially inducible mouse models reveal that Npm1 mutation causes malignant transformation of Dnmt3a-mutant clonal hematopoiesis. *Leukemia* 1 (2019). doi:10.1038/s41375-018-0368-6
43. Meyer, S. E. *et al.* DNMT3A Haploinsufficiency Transforms FLT3 ITD Myeloproliferative Disease into a Rapid, Spontaneous, and Fully Penetrant Acute Myeloid Leukemia. *Cancer Discov.* **6**, 501–15 (2016).
44. Nangalia, J. *et al.* Somatic *CALR* Mutations in Myeloproliferative Neoplasms with Nonmutated *JAK2*. *N. Engl. J. Med.* **369**, 2391–2405 (2013).
45. Campbell, P. J. & Green, A. R. The Myeloproliferative Disorders. *N. Engl. J. Med.* **355**, 2452–2466 (2006).
46. Jaiswal, S. *et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
47. McKerrell, T. *et al.* Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Rep.* **10**, 1239–45 (2015).
48. Shepherd, M. S. *et al.* Single-cell approaches identify the molecular network driving malignant hematopoietic stem cell self-renewal. *Blood* **132, (2018).

Using of single-cell approaches alongside combinations of mutations to allow the study of the drivers of individual disease characteristics - in this case self-renewal in myeloproliferative neoplasms.

49. Papaemmanuil, E. *et al.* Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**, 3616–27; quiz 3699 (2013).
50. Ortmann, C. A. *et al.* Effect of Mutation Order on Myeloproliferative Neoplasms. *N. Engl. J. Med.* **372**, 601–612 (2015).
51. Leal, A. D. *et al.* Anthropometric, medical history and lifestyle risk factors for myeloproliferative neoplasms in the Iowa Women's Health Study cohort. *Int. J. cancer* **134**, 1741–50 (2014).

52. Schuettpelz, L. G. & Link, D. C. Regulation of Hematopoietic Stem Cell Activity by Inflammation. *Front. Immunol.* **4**, 204 (2013).
53. King, K. Y. & Goodell, M. A. Inflammatory modulation of HSCs: viewing the HSC as a foundation for the immune response. *Nat. Rev. Immunol.* **11**, 685–692 (2011).
54. Hawkins, E. D. *et al.* T-cell acute leukaemia exhibits dynamic interactions with bone marrow microenvironments. *Nature* **538**, 518–522 (2016).
55. Petti, A. A. *et al.* Mutation detection in thousands of acute myeloid leukemia cells using single cell RNA-sequencing. *bioRxiv* 434746 (2018). doi:10.1101/434746
* 56. Rodriguez-Meira, A. *et al.* Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *Mol. Cell* (2019). doi:10.1016/J.MOLCEL.2019.01.009

The next step following on from Giustacchini et al (2), allowing combining scRNAseq with genotyping from with cDNA and gDNA allowing many more mutations to be identified in single cells - future application of this technology could provide insights into many diseases.

57. Pascutti, M. F., Erkelens, M. N. & Nolte, M. A. Impact of Viral Infections on Hematopoiesis: From Beneficial to Detrimental Effects on Bone Marrow Output. *Front. Immunol.* **7**, 364 (2016).
58. Matatall, K. A. *et al.* Chronic Infection Depletes Hematopoietic Stem Cells through Stress-Induced Terminal Differentiation. *Cell Rep.* **17**, 2584–2595 (2016).
59. Takizawa, H. *et al.* Pathogen-Induced TLR4-TRIF Innate Immune Signaling in Hematopoietic Stem Cells Promotes Proliferation but Reduces Competitive Fitness. *Cell Stem Cell* **21**, 225–240.e5 (2017).
60. Feng, C. G., Weksberg, D. C., Taylor, G. A., Sher, A. & Goodell, M. A. The p47 GTPase Lrg-47 (Irgm1) Links Host Defense and Hematopoietic Stem Cell Proliferation. *Cell Stem Cell* **2**, 83–89 (2008).
* 61. Lee-Six, H. *et al.* Population dynamics of normal human blood inferred from somatic mutations. *Nature* **561**, (2018).

Using somatic mutations to trace a clonal history HSCs, allowing inferences about the population dynamics of HSCs - including population size, time between self-renewal divisions and the dynamics of differentiated cell output over time. Future work on the effect of mutations on clonal dynamics are readily anticipated.

62. Kirschner, K. *et al.* Proliferation Drives Aging-Related Functional Decline in a Subpopulation of the Hematopoietic Stem Cell Compartment. *Cell Rep.* **19**, 1503–1511 (2017).
63. Loughran, S. J. *et al.* Mbd3/NuRD controls lymphoid cell fate and inhibits

tumorigenesis by repressing a B cell transcriptional program. *J. Exp. Med.* **214**, 3085–3104 (2017).

* 64. Hérault, A. *et al.* Myeloid progenitor cluster formation drives emergency and leukaemic myelopoiesis. *Nature* **544**, 53–58 (2017).

Study on the spacial organisation of progenitors within the bone marrow (GMPs, under normal homeostatic conditions, cancer and stress) and the gene expression patterns that accompany different spacial organisations.

65. Guo, G. *et al.* Mapping Cellular Hierarchy by Single-Cell Analysis of the Cell Surface Repertoire. *Cell Stem Cell* **13**, 492–505 (2013).

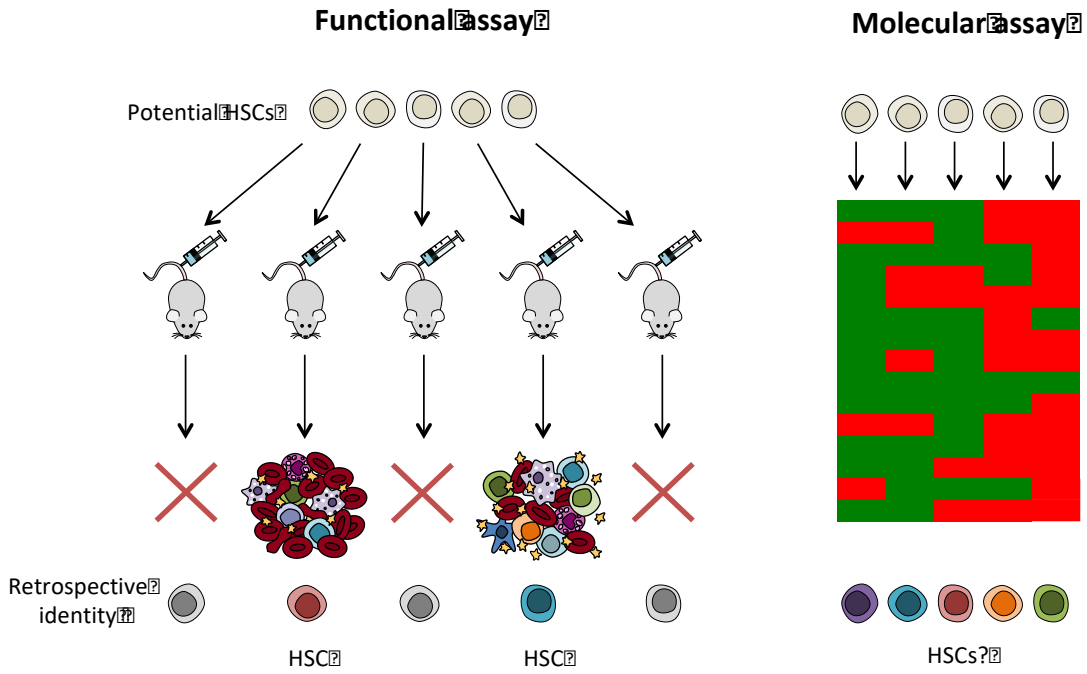

Figure 1: Selectivity of functional assays reduces experimental noise in the description of HSC heterogeneity. Depicted on the left is a single cell functional assay where 5 input cells are each transplanted into a single recipient mouse and analysed for the presence of an HSC. In this case, 3 cells do not read out and a "red" HSC subtype and a "blue" HSC subtype are easily compared. On the right are a similar set of 5 input cells where each cell is subjected to single cell gene expression profiling – all 5 cells generate a molecular profile that is unique, but it is unclear which profiles belong to HSCs and which belong to non-HSCs, making subtype classification more difficult.

Ref	Mutation	Cell type	SC Molecular Techniques	SC Functional Techniques	Key conclusions
			Single Mutant		
Kirschner et al 2017 Cell reports[62]	JAK2	ESLAM HSC	scRNA-seq	Sc HSC liquid culture	Young vs old JAK2 V617F HSCs; in aged HSCs increased p53, accompanied by functional decline
Loughran et al 2017 JEM [63]	Mbd3	HSCs, LMPPs, ALPs, and BLPs	Sc-qPCR & scRNA-seq	Sc lymphocyte differentiation culture (OP9/ OP9 DL1)	Mbd3/NuRD complex represses B cell transcriptional program, preventing lymphoid progenitors from undergoing B cell lineage commitment
Herault et al 2017 Nature[64]*****	BCR-ABL	GMP	scRNA-seq & sc-qPCR	Sc CFCs, in situ imaging	Spatial organisation of GMPs; clusters formed in emergency myelopoesis and BCR-ABL mutant have enhanced self renewal networks
Guo et al 2013 Cell Stem Cell[65]	MLL-AF9	LSK, LSK^-, LS^-K, LS^-K^-	Sc-qPCR	None	Leukemia cells are intrinsically distinct from any wild-type lineages, Ezh2 is overexpressed in the more highly proliferative leukemia cells
			Double Mutant		
Meyer et al 2016 Cancer discovery[43]	DNMT3a FLT3ITD	c-kit AML splenocytes	ScRNA-seq	None	Investigate whether c-Myc expression could be a biomarker for $FLT3^{ITD}$
Shepherd et al 2018 Blood[48]	JAK2 TET2	ESLAM HSC	Sc-qPCR	Sc HSC liquid culture	Molecular regulators of JAK2 V617F stem cell defect identified as *Bmi1, Meis1,* and*, Pbx1*.

Table 1: Table of studies which have used single-cell molecular techniques to study mouse models of haematological malignancies.